Financial Incentives in State Accountability Systems:
Performance Pay for Teachers

Carolyn Kelley
University of Wisconsin-Madison
Consortium for Policy Research in Education
1025 West Johnson Street, Rm 1161G
Madison, WI 53706
608-263-5733 (phone)
608-265-3135 (fax)

September 10, 2002

This paper was prepared for the Center on Reinventing Public Education at the University of Washington and the Washington State Academic Achievement and Accountability Commission. Some of the research reported in this paper was supported by a grant from the U.S. Department of Education, Office of Educational Research and Improvement, National Institute on Educational Governance, Finance, Policy-Making and Management, and the Pew Charitable Trusts (No. 97001184000) to the Consortium for Policy Research in Education and the Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison (Grant No. OERI-R3086A60003). The opinions expressed are those of the author and do not necessarily reflect the views of these institutional partners.

Financial Incentives in State Accountability Systems:
Performance Pay for Teachers

By 2002, 32 states provided rewards for the achievement of educational goals and/or sanctioned schools failing to achieve goals (ECS, 2002). Typically, reward plans pay bonuses to all of the teachers in a school, to teachers and school staff members, or provide one-time discretionary funds to schools to reward employees for meeting school performance goals. The establishment of high-level expectations for all students and the Federal No Child Left Behind Legislation provide added impetus for states to consider the use of financial incentives to focus and direct teacher effort.

School-based performance award (SBPA) programs are modeled after compensation systems developed in team-based organizations in the private sector. These pay plans attempt to better align compensation incentives with the achievement of organizational goals. Since goal achievement in team-based organizations requires that employees work together to meet group goals, compensation systems have been modified to include rewards for achievement of group performance objectives (Lawler, 1990; Schuster & Zingheim, 1992).

Team-based, collaborative work styles are also a natural fit for schools (Odden & Kelley, 2002). Like accountability systems generally, pay for performance systems provide guidance to teachers about the expected and desirable goals. They also provide important symbolic and substantive guidance to school and district administrators regarding important educational and policy priorities of the state. Thus, performance pay systems provide direction to teachers to focus their efforts, including aligning curriculum and individual course content to state standards, aligning teaching practice to the test, and focusing individual and collective professional growth toward improving educational outcomes on the rewarded goals.

SBPA programs use a portion of teacher pay to reward group goals. Typically, the bulk of pay continues to be distributed using the traditional single salary schedule, which pays teachers for years of experience and the accumulation of educational credits and degrees. SBPA programs differ from traditional pay for performance (or "merit pay") plans that focus on identifying and rewarding a limited number of teachers in a school for performance excellence. Often characterized by unclear performance objectives, limited observation of teaching, and quotas on the number of teachers who could be identified as excellent, these merit pay plans were divisive, arbitrary, and short-lived (Hatry, Greiner, & Ashford, 1994; Murnane & Cohen, 1986). They also failed to support collaboration among teachers, which researchers have found to be essential to high performance in schools.

The design of school-based performance pay systems are critical to their success, as what appear to be minor differences in design choices can have profound effects on the legal defensibility of the system, the ability of schools to achieve the goals and be

rewarded, and the influence on teacher behavior and teacher and student performance. Symbolically the systems can be powerful, affecting the perception of school quality in the community, providing important incentives for schools to improve, and affecting the career choices and options of teachers and school and district administrators.

In the short term, performance pay may influence school performance by affecting teacher effort in at least 3 ways: by affecting the *focus* of teacher effort; by affecting the *amount* of teacher effort; and by affecting the *quality* of teacher effort. Without systemic realignment and investment in teacher knowledge and skills and improvements in organizational performance, SBPA programs work only to the extent that teachers already have the capacity to change practice to improve student performance. Thus, typically, short-term gains in student performance reflect stepped-up efforts to align instructional content with state standards, and efforts to prepare students for the test by familiarizing them with its approach and hyping its importance.

Ongoing efforts to improve performance can be more difficult, and require investment of new resources and alignment of existing resources to develop teacher knowledge and skills needed to achieve long-term performance goals. Achieving long-term gains in student performance requires significant and substantial change to teaching practice, and therefore must be undertaken with significant investment in and an understanding of the ways that school organizations need to change to become more effective (Elmore, 2002).

In theory, school-based performance pay provides a way to cut through the multiple, competing goals of schools, to provide clear performance objectives for teachers to work toward. By allocating a small portion of pay to reward group performance, school-based performance awards encourage collaboration among teachers. This corrects an important problem created by individual performance pay plans that discouraged collaborative problem solving among teachers.

The Effects of School-Based Performance Pay

There is a growing body of evidence about the effectiveness of school-based performance award programs. For the most part, the evidence is consistent with expectations. Research shows that SBPA programs are associated with:

- clear goals for teachers, administrators, and policymakers;
- alignment of resources to support these goals;
- improvements in student achievement, particularly on the rewarded goals (http://www.wcer.wisc.edu/cpre/; Milanowski, 1999; Pogglio, 2000, Smith, Rothacherand & Griffin, 1999); and
- outcomes many teachers view as positive, including the monetary bonus, opportunities to see student performance improve, opportunities to collaborate with other teachers, and opportunities to participate in meaningful professional development (Heneman & Milanowski, 1999; Kelley, Heneman & Milanowski, 2002; Kelley & Protsik; 1997).

The research suggests that program design and implementation processes are critical to the success of school-based performance pay (Hatry, Greiner & Ashford, 1994; Leithwood, 2000; Odden & Kelley, 2002).

Existing evidence suggests that SBPA programs are associated with improvements in student performance. For example, Poggio (2000) analyzed performance data in Kentucky between 1992 and 1998, and found that student test scores rose in all subjects, with largest gains occurring at the elementary and high school levels. Typically, the largest gains occurred in the first two years of the program, with more modest and erratic gains in later years. Table 1 summarizes performance gains in the reading, math, science and social studies between 1992 and 1998. The scores represent significant increases in performance. The goal of the Kentucky program was to move every school to a score of 100, or proficient, in all subjects over a 20-year period. In 1998, the state changed assessment instruments, making comparisons across time more complicated.

Table 1. Accountability Index Scores by Subject for Kentucky Schools, 1992-98

	Reading		Math		Science		Social Studies	
School Level	1992-3	1997-8	1992-3	1997-8	1992-3	1997-8	1992-3	1997-8
Elementary	32.4	58.4	22.3	44.4	18.2	37.2	27.6	37.9
Middle	38.4	47.0	22.8	51.4	19.9	21.7	28.9	33.0
High	20.2	51.4	22.2	47.0	27.1	44.3	23.4	49.4

Source: Poggio, 2000

Milanowski (1999) has also examined achievement gains in Kentucky schools and found that the gains are not simply attributable to random variation. Kane & Staiger did a related analysis on North Carolina data and found that while a significant proportion of changes in test scores are due to random variation, some of the increase from year to year can be attributable to real gains in student performance (Kane & Staiger, 2001).

Similarly, in North Carolina, the number of students at or above grade level in reading and math has increased steadily over the past decade. Following implementation of the ABCs accountability model, there was a slight increase in the upward sloping performance trend. Since 1996-7 when the ABCs program was implemented, the percentage of students performing at or above grade level in reading and math, grades 3 through 8 increased from 61.7 to 71.7 percent (up from 52.9% in 1992-93). The performance gains have been consistent for all racial subpopulations of students; as a result, the gap between African-American (the lowest performing group) and White students (the highest performing group) has remained a relatively consistent and troubling 30 percentage points. In 2000-01, African American students scored at 52% at or above grade level in reading and math, while White students scored at 82% at or above grade level in these subjects (North Carolina Department of Public Instruction, 2001a).

Similar results have been obtained at the district level in Dallas (Texas) following the implementation of school-based performance award programs in that district. In analyses comparing Dallas to similar-sized Texas districts, Clotfelter and Ladd found that

performance gains in Dallas out-paced gains in similarly situated urban Texas districts (Clotfelter & Ladd, 1996; Ladd, 1999). Of some concern, the Dallas study found gains for White and Hispanic students, but not for African American students. Ladd (1999) also found higher principal turnover rates in Dallas than in the comparison districts.

While the performance gains found in these studies of Kentucky, North Carolina, and Dallas were the result of a combination of educational policy initiatives, they provide circumstantial evidence that the performance pay plan may have a positive impact on student performance.

Performance improvement is most evident as measured by the assessment instruments and measures rewarded by the programs, and somewhat less evident using other measures (Klein, Hamilton, McCaffrey & Stecher, 2000; Poggio, 2000). The failure to improve significantly on non-rewarded instruments may be due to narrow teaching to the test, the higher stakes of the rewarded test, or the insensitivity of test instruments to measure changes in learning gains (Baker, 2002).

SBPA programs may create conditions for the improvement of student performance in least three ways. First, the program establishes clear instructional goals that teachers and students can work toward. Thus, some improvement may occur as a result of familiarizing students with the assessment process, and making sure that the curriculum content on the assessment is covered in the curriculum.

Second, SBPA programs provide incentives for teachers to encourage students to do their best on the assessment. Strategies employed at schools in Kentucky include invoking school pride, providing a positive test-taking atmosphere, making the test high stakes for students, and rewarding students for their efforts with prizes and parties after the exam (Kelley & Protsik, 1997).

Third, over time, teachers and administrators have an incentive to implement improvement strategies to meet program goals. These efforts may involve both alignment and improvement of curriculum and instructional approaches, and can include:

- examining prior test scores, and revising teaching practice to address areas of identified weakness;
- strategically aligning curriculum within and across grade levels with the standards and assessments;
- developing new curricular materials and instructional approaches;
- investing in professional development related to the goals of the assessment; and
- collaborating with other teachers to improve performance (Kannapel et al., 2000; Kelley, 1998; Kelley & Protsik, 1997; Stecher & Barron, 1999).

Some researchers and practitioners have expressed concern that SBPA programs may overly narrow the curriculum, divert it from important instructional goals that are not assessed, and warp educator behaviors (Kannapel et al., 2000; King & Mathers, 1997; Pogglio, 2000; Stecher & Barron, 1999; Stecher & Hamilton, 2002). This “teaching to

the test" creates potential measurement problems in meaningfully assessing student learning. Most tests estimate student learning by sampling student knowledge. But if the curriculum is carefully aligned to the assessment, rather than representing a sampling of student work, the test may measure a narrow knowledge base developed solely to improve test performance, rather than to achieve broader learning goals (Klein et al., 2000; Pogglio, 2000).

The problem may be exacerbated by narrow or low-level assessments, and by programs that provide inordinately large bonuses for performance improvement. Research findings from case studies conducted by the Consortium for Policy Research in Education (CPRE) are consistent with research in the private sector, which suggests that bonuses of about $2000 to $3000 are probably sufficient to influence teacher behavior without triggering significant levels of gaming behavior (Kelley, Heneman & Milanowski, 2002).

Some states have taken care to design assessments that promote the kinds of teaching practices desired. In effect, these assessments are designed to be "taught to." But more often, the assessments used are off-the-shelf, and of variable quality. Certainly low-level basic skills assessments may move curriculum away from challenging content and pedagogical practices toward drill and practice, watering down the curriculum. Test developers continue to strive to develop assessments that can successfully drive excellent teaching practice (Koretz & Barron, 1998).

It is also important to note that these educational policy interventions are put in place in dynamic policy contexts, so there is no "pure" experiment to prove the effectiveness of SBPA programs. However, the evidence suggests that teachers working under SBPA programs do change their instructional practices to align them more closely with state learning and assessment goals (Kannapel et al., 2000; Kelley, 1998; Stecher & Barron, 1999).

How do SBPA Programs Work?

Studies of SBPA programs conducted by the Consortium for Policy Research in Education suggest that schools were more likely to achieve their goals when teachers believed that their effort would lead to increases in school-wide performance (Kelley, Heneman & Milanowski, 2002). In the psychological literature, this is referred to as *teacher expectancy*. Factors that enhanced teacher expectancy, and thereby enhanced school success at meeting performance achievement goals, included the presence of effective feedback mechanisms, a sense of professional community or leadership, a history of success with the award program, teacher perceptions that the program was fair, and a lack of conflict between school and program goals (Kelley, Heneman & Milanowski, 2002).

The performance award programs also appear to encourage collaboration. Among the outcomes teachers identified as positive consequences of the award program, teachers identified working with colleagues (Heneman & Milanowski, 1999) as an important positive effect. Further, an evaluation of the Douglas County Performance

Award Program found similar results. School-based performance pay was one of eight elements of a compensation plan in Douglas County. Teachers reported that this portion of the plan was the most popular element of the pay plan as a whole (and the plan was in a contract approved by teachers with almost unanimous support); and teachers indicated that it promoted significant collaboration among teachers to achieve the goals of the plan (Hall & Caffarella, 1997).

Teacher awareness of, and teacher motivation under, school-based performance pay appears to be stronger when the money is used for teacher salaries rather than as additional discretionary money in the school budget. Teachers were less familiar with, and seemed to be somewhat less affected by the accountability program in Maryland (that rewarded schools but did not allow the money to be used for pay bonuses) than they were in Kentucky (which allowed the money to be used for pay bonuses). While it is difficult to separate the effects of state context from the specific program design effects, it does appear that program design had something to do with lower levels of awareness and salience of the program in Maryland (Kelley, Conley & Kimball, 2000).

Other Considerations

In addition to affects on teaching practice, SBPA programs can affect employee morale. While it is difficult to disentangle the effects of rewards versus sanctions, since they are typically present together in most programs, accountability programs linked to school performance do appear to heighten teacher anxiety and teacher workload (Heneman & Milanowski, 1999; Kelley, 1998). The negative effects of SBPA programs appear to be lessened or alleviated when important enabling conditions are in place, such as:

- principal leadership, determination, and support of teachers to achieve program goals;
- teachers own beliefs that they can achieve program goals; and
- the presence of administrative supports, such as effective, focused professional development and opportunities to evaluate feedback data, with training in how to use this information to improve practice (Kelley, 1998).

The research on SBPA programs overall suggests that program goals should be set at reasonable levels, and important organizational and policy supports should be in place to help teachers achieve program goals (Kelley, Heneman & Milanowski, 2002). Some have argued that a complementary policy of knowledge and skills based pay should also be in place, to reward individual teachers for the development of knowledge and skills needed to achieve SBPA program goals (Mohrman, Mohrman & Odden, 1996; Odden & Kelley, 2002).

Cautions

Compensation is not a blunt policy instrument. To be effective, pay systems must be designed carefully, with attention to the assessment instruments used to measure performance improvements, the level of goals established, and the support teachers have to modify instructional strategies to achieve program goals. When teachers are put under

tremendous pressure to improve without the resources needed to make the improvements, SBPA programs can destroy morale and warp educational objectives.

Teachers need to understand the goals, see them as achievable, feel supported in improvement efforts, and trust that the system will be carried out fairly so they will receive the promised rewards. While SBPA programs can narrow the curriculum, they can also provide goal focus and data feedback loops for teachers to evaluate and improve their own teaching practice. Thus, school-based performance pay can be an important component of an overall policy strategy to support teachers in improving student performance.

The No Child Left Behind legislation raises the stakes for states, school districts, and teachers generally to show improvements on state and national assessments. Pay for performance is one mechanism that may support teacher efforts to improve student performance on these measures.

Examples of State Reforms

The following section provides an overview of state pay for performance plans operating in North Carolina, Kentucky, California, Florida, and Iowa. These plans provide an overview of some of the issues that arise in developing and implementing performance pay plans, and provide examples of features typical of the more carefully designed programs. They also illustrate some of the challenges in designing and implementing incentive programs.

Of the five, North Carolina and Kentucky are examples of programs that have undertaken comprehensive and systemic reform strategies that include financial incentives as a centerpiece of accountability. California is an example of a state that tried to leverage significant financial incentives to influence teacher behaviors and teacher employment markets. Florida has implemented a system of rewards, sanctions, mandates and supports in an attempt to mobilize schools for significant educational reform. And Iowa is a unique example of a state that is attempting to use compensation to leverage improvements in teacher quality and capacity.

The North Carolina and Kentucky experiences are interesting because they represent significant new investment and a comprehensive reform strategy that makes rewards a part of a large scale effort to significantly improve student learning. These states have experienced noticeable improvements in student performance over the life of the performance pay plan. They have come as close as any state to "doing it right" – through:

- careful selection and continued investment in assessment instruments,
- attention to the impact of the assessments on teaching,
- public engagement to explain the assessment and results,
- a willingness and commitment to ongoing evaluation and revision of the plan to enhance its effectiveness, and

- a significant commitment to an overall reform strategy that addresses conditions required to enable teachers and schools to succeed at improving student performance.

California implemented a series of financial incentives, the bulk of which were in place for only one year. The results were chaotic, perhaps due to features of the California program design, which included very large rewards distributed based on a quota system. The California plan was also the product of incremental politics more than a comprehensive, carefully designed reform plan that links changes in pay to an overall education reform strategy. The pay incentives were implemented in a system suffering from inadequate support for public education, and a crisis in teacher quality and supply created by an educational reform strategy of reducing class sizes on a massive scale.

The Florida experience is interesting as well, because it combines a strong data-driven assistance strategy with a high profile public reporting system that has significant rewards and sanctions attached to it. All of these systems use improvement in performance or a combination of performance improvement and high levels of performance as the criteria for measuring goal achievement.

The last example, Iowa, is the only state thus far that has attempted to move toward replacing the traditional teacher salary schedule grounded in years of experience and educational credits and degrees with one that rewards teachers more directly for the development and demonstration of knowledge and skills as a result of a perceived crisis in teacher quality and supply in that state.

North Carolina

North Carolina's school-based performance award program is called the ABC's of Public Education: A – **Accountability** for educational standards; B – emphasis on the **Basics**; and C – maximum local **Control**. The plan was piloted in 1995-96, implemented in elementary and middle schools in 1996-97, and in high schools in 1997-98. Adjustments have been made to the plan annually since its implementation.

The ABC program sets growth and performance targets individually for each elementary, middle and high school in the state based on absolute performance standards (the percent of students scoring at or above grade level in a school) and relative growth standards. Expected growth for a school is based on previous performance, statewide average growth rates, and a statistical adjustment for comparing student scores from one year to the next (regression to the mean). While the formula for determining growth targets is quite complex, each school receives a software program that enables them to calculate their own school's expected and actual gains according to the ABCs model.

The ABC accountability plan includes accountability, recognition, assistance, and intervention. The state rewards teachers in schools that reach expected or exemplary (above expected) growth with $1500 per teacher. Teachers in schools that reach expected growth receive $750 per teacher. The state also publicly recognizes schools that achieve expected and exemplary growth, with additional recognition for the top schools across

the state in the level and growth in performance. In 2000-01, 768 schools met their expected growth targets and received awards (North Carolina Department of Public Instruction, 2001).

Schools with overall low performance that fail to achieve growth targets over time are designated low-performing schools, and are assigned an Assistance Team. Assistance Team members include "currently practicing teachers and state representatives of higher education, school administrators, retired educators, and others the SBE considers appropriate" (North Carolina Department of Public Instruction, 2002). The Assistance Team reviews school and district administrator and teacher performance, as well as overall school performance, collaborates in the design, implementation and monitoring of a school improvement plan, reviews and reports on school progress, and can make recommendations to the State Board of Education on termination of individual teachers or administrators, including the school principal and district administrator. In 2000-01, 31 schools were identified as low performing, and of these, 14 schools received assistance teams. The number of low performing schools has declined each year of the program, and is down from 7.1% of all schools in the first year of the program (1996-97) to 1.4% of schools in 2000-01 (North Carolina Department of Public Instruction, 2001).

In addition, the state has established student accountability standards that require that students perform at grade level in order to be promoted at the 3rd, 5th, 8th, and 11th (for graduation) grade levels. The 11th grade test is being implemented in Spring 2004 for the graduating class of 2005. The state has made accommodations and exemptions for students to enable alternative assessment through a year-long portfolio process, accommodation for special needs students and exemptions for up to 2 years for Limited English Proficient students.

The state has maintained a commitment to continuously reviewing and revising the program over time to address problems as they emerge. For example, the DPI notes that "Every year since the implementation of the ABCs there have been a few schools where a small number of students in a subject or grade have had unduly positive or negative influence on the ABCs growth of a school. Weighting of the ABCs growth composites was adopted. . . to deal with such disproportional impacts" (NCDPI, 2002, p. 17). Further, the state reviews the ABCs model annually, including through the use of simulation models, to ensure that the model remains stable as components are adjusted over time.

The state has identified characteristics of the most improved schools, which include the review and use of student test data to develop instructional and curricular goals, strong leadership, goal focus, stability in staffing, ongoing technical support and assistance, parental involvement and responsible decision-making (NCDPI, 2002, p. 16). They have also seen instructional improvements over time in low-performing schools, as well as increased familiarity in these schools with the state content standards.

Among the challenges identified by the Department with respect to program implementation include significant costs associated with collecting, analyzing, and

reporting test results to schools and communities; effective communication about the meaning of assessment results given the complexity of the growth formula calculations; significant investment in studies to ensure consistency of calculations of learning growth over time as new assessment come on line and/or are incorporated into the ABC formula; managing teacher perceptions about the test (e.g., teacher concerns about teaching to the test, narrowing the curriculum, spending too much time being tested or practicing for tests) by encouraging teachers to focus on aligning instruction to the state content standards and by carefully managing the use of language to describe the program (e.g., balancing recognition for high achievement with recognition for growth in achievement).

The effects of the program were summarized by the state Department of Public Instruction as follows:

> North Carolina's ABCs of Public Instruction has made a positive difference in performance of students and professional educators across the state. By using this accountability model, we have found that successful schools focus on the *Standard Course of Study*, use data-driven decision making, align curriculum and instruction with testing, and in turn have increased student achievement (NCDPI, 2002, p. 19).

North Carolina has also made a significant investment in the professionalization of teachers and in the development and recognition of teacher knowledge and skills. At 12% of salary, the state provides one of the largest and longest-running pay rewards to teachers who achieve National Board Certification (NCDPI, 2002).

Kentucky

Kentucky has had a school-based performance awards plan in place since 1993-94. Like North Carolina, Kentucky's plan reflects a mature system that has had time to identify and correct emerging and recurring problems. And like North Carolina, Kentucky has invested significant resources in the development of the assessment and reporting systems, in efforts to appropriately identify performance levels and gains, and in developing an intervention system that identifies the approaches and behaviors of teachers and leaders in successful schools, and assigns intervention assistance to schools that fail to improve.

The SBPA plan was adopted as part of Kentucky's 1989 court-ordered overhaul of the education system. As with North Carolina, performance awards in Kentucky are part of a systemic reform strategy to align standards, assessments, professional development, rewards and sanctions, and educational intervention strategies with a focus on improving student performance across the state.

Kentucky's standards, assessment and accountability system was developed with an eye toward transforming the content and pedagogy used by teachers to enhance learning outcomes. Initially, the state developed and adopted a student assessment system called KIRIS, the Kentucky Instructional Results Information System. The KIRIS system lost political favor due to concerns about its validity and reliability, and in

1998 the General Assembly required that it be replaced by a new system, the Commonwealth Accountability Testing System, or CATS. Over time, the Kentucky system has seen substantive and symbolic changes to refine, improve and protect the system from political attack.

The assessment itself covers reading, math, language arts, writing, science, social studies, arts and humanities, and practical living/vocational education measured through a combination of multiple choice, open response, and portfolio entries. For the purposes of accountability, students' performance is measured in the 4th and 5th, 7th and 8th, and 11th and 12th grades. Student performance is assessed across cohorts, although a longitudinal component is being developed that could show progress of the same students over time.

Student performance is assessed at four levels: novice, apprentice, proficient, and distinguished. Beginning in 1999, the categories were expanded from 4 to 8 in reading, math, science and social studies in order to give schools credit for growth in student performance that did not cross the broader thresholds in the four categories (e.g., some schools expressed concern that student scores grew over time, but not enough to move them to the next category, and the older system did not recognize this growth).

The system has an overall goal of all schools achieving a score of 100 (or "proficient") by 2014. This fall, each school will receive its own growth chart, showing progress needed to reach 100 by 2014. The growth chart will appear in each school report card. Schools are rated based on their growth in performance compared to their progress toward the goal as follows: Meeting goal (improving at a rate needed to achieve 100 by 2014); Progressing (scoring above 80 or improving but at a rate below the goal rate); or Assistance (dropping below the baseline established as the average of the 1998-99 and 1999-00 school year scores).[1]

Schools have four ways to achieve awards:

- Schools that meet their goals, satisfy the drop-out requirement and continually reduce the number of students in the novice category;
- Progressing schools that score higher than in the previous 2-year period, have met the goal for reducing the number of students in the novice category, and meet drop-out goals receive a half-share of the rewards;
- Schools improving enough to meet certain score levels receive awards; and
- Pacesetter Schools, those performing in the top 5 percent of all schools, have an acceptable dropout rate, and not declined in the past 2 cycles, and have not received other CATS awards receive rewards.

[1] These labels have been modified over time to provide clearer information to the public. Initially, schools were identified as "reward," "successful," "improving," "in decline," and "in crisis." The state had a difficult time communicating the meaning of these labels to the public, particularly since very high performing schools could be labeled "in decline" if they were not improving their performance, and very low performing schools could be labeled "successful" or "reward" schools if they showed improvements in student performance.

Rewards are paid to the school and the site council (or principal if there is no site council) determines their use, which can include school uses or staff bonuses.

Depending on their place in the distribution, schools falling in the assistance category have varying levels of intervention required. In the top third, assistance schools must conduct a self-study, may apply for improvement funds, and may recommend changes to the teacher evaluation process. In the middle third, assistance schools must conduct a self-study, may apply for improvement funds, and must recommend changes to the teacher evaluation system. In the bottom third, assistance schools are assigned a highly skilled educator[2], are subject to a scholastic audit conducted by a state team, may apply for school improvement funds, and are subject to teacher evaluation review, recommendations for changes to professional growth plans, and evaluations that could lead to principal, teacher, and school council member dismissal. Students in schools that remain at level 3 for two or more years may transfer at district expense to a successful school (Prichard Committee, 2001).

In 1999-2000, the most recent award cycle, 618 schools improved enough to receive a share of $23 million in rewards. 149 schools fell below expected performance and were eligible for assistance; among these 49 schools fell in the bottom third of the distribution, and received a scholastic audit (Kentucky Department of Education, 2000). Results for the next biennium will be released in Fall, 2002. In 2001, Kentucky completed a process of developing new performance standards to provide clearer information to Kentucky teachers about content mastery required for students to achieve at the novice, apprentice, proficient, and distinguished levels (Wilhoit, 2001).

The Kentucky reforms are also characterized by a significant investment in educating and engaging the public to understand and provide ongoing support for education reform in Kentucky. The work of Prichard Committee and others to engage in an ongoing large-scale effort to educate and inform the public has helped to maintain public support to sustain the reform effort, and continues to generate both support and constructive pressure to improve public education in Kentucky.

The evolution of the Kentucky performance pay system illustrates the importance of an emphasis on teacher knowledge and skills. As the system placed pressure on schools to improve, it also highlighted professional growth needs of teachers to achieve

[2] Highly Skilled Educators are Kentucky educators selected and trained by the state to provide assistance to schools failing to show improvements in student performance. The Highly Skilled Educators and their predecessors, the Distinguished Educators, have a proven track record in their ability to assist schools in improving student performance. The ability of schools to sustain this improvement depends in part on their efforts to institutionalize the reforms, such as through hiring a school-site curriculum/instructional improvement coordinator. Highly Skilled Educators represent an internal assistance model which avoids some of the problems inherent in a decentralized external provider approach such as the one adopted to assist low performing schools in Chicago. The Highly Skilled Educators provide an external reviewer for school processes and approaches, coordinated at the state level, but they are insiders to the extent that they have had successful experiences in similarly situated schools from the same region of Kentucky (Wakelyn, 2002 and personal conversation with D. Wakelyn, June 3, 2002).

these significantly higher expectations for student learning. This year, the state legislature passed legislation that provides "flexibility for school districts to provide added compensation above the single salary schedule for certain purposes, including rewards for teachers who increase their skills, knowledge and instructional leadership. The law establishes a fund to provide grants to at least five school districts that will conduct pilot studies of differentiated compensation programs" (Salyers, 2002, p. 1). This reflects ongoing interest by Kentucky policymakers in developing workable compensation plans that reward the development of key teacher knowledge and skills.

Kentucky has also recently added a bonus of $2000 per year for teachers achieving National Board Certification status, and has a goal of at least one National Board Certified teacher in each school by 2020 (Salyers, 2001)

California

In 1999, the California State Legislature adopted the Public Schools Accountability Awards Act, which provided funds to certificated staff for schools that exceeded performance targets and had performance in the top 50% of the statewide distribution as measured by student test scores. The pay bonus program provided bonuses of $5000, $10,000 or $25,000 to the schools with the largest one-year growth in student performance in the state (California Department of Education, 2002). Funding was eliminated after one year due to a combination of the state's significant budget shortfall and organized teacher opposition to the program.

The state continues to provide a $10,000 bonus to teachers achieving National Board Certification. National Board Certified teachers who choose to teach in a low performing school (defined as below the 50th percentile in state student test scores) can receive a $25,000 bonus. The state thus far has been unable to track how many National Board Certified teachers in low performing schools moved to the school after becoming Board Certified, compared to the number of teachers who achieved Board Certification that were already in these low performing schools. State funds to evaluate the Board Certification bonus program were eliminated due to the budget shortfall, so the results of this part of the program have yet to be evaluated.

The California program illustrates some potentially problematic design features of school-based performance award programs, including:

- the large size of the bonus;
- the quota style system for determining who would receive the large awards;
- the failure by the state to maintain funding for the program over time to ensure that the program would have incentive value in the future; and
- the incremental nature of the reform, which was not implemented within an organized and systemic reform strategy.

As a result, the program produced an environment ripe for manipulation, with some schools dissuading significant numbers of students from taking the test, prompting the state to add rules post-hoc requiring reward schools to have at least 95% of students

taking the test in elementary and middle schools, and 90% in high schools. The quota system created a situation in which schools didn't know whether they would receive the $25,000, $10,000 or $5,000 bonus until well into the next school year. The measurement of growth was based on one year of data, compounding the likelihood of schools becoming eligible for awards due to chance fluctuations in student performance on the assessment. Because the bonus award was abandoned after one year, there was insufficient opportunity for most schools to respond to the award incentive by making meaningful changes in the organization and practice of teaching (LAO, 2002). This further contributes to lottery-style perceptions of "winning" the award. Because the bonus was withdrawn after one year, future efforts to use financial incentives to motivate changes in teaching practice are likely to be largely unsuccessful in California.

Furthermore, some have criticized the state's programs for putting too much emphasis on standardized test scores that were not sufficiently aligned with the state's academic standards. These criticisms have contributed to the state's recent decision to work with the Education Testing Service (ETS) to develop standards-based assessments in English, math, history/social studies, and science. ETS is also currently working on developing a High School Exit Exam for California. The new standards-based exams and the California Achievement Test would replace the SAT-9 and a standards-based English/language arts assessment that were the basis for the calculation of the Academic Performance Index (API) that determined school eligibility for awards in 2000-2001 (Olson, 2002).

Florida

The Florida School Recognition Program has been in place for five years. The plan includes grading of schools from "A" to "F," based the level of student performance on the state's criterion referenced exam, improvement in individual student performance from the previous year, and improvements in the performance of the lowest 25% of students in reading. "A" schools must have at least 95% of students taking the test; "B" through "F" schools are required to test at least 90% of their students. Students are tested in grades 3 through 10 in reading and math, and grades 4, 8, and 10 in writing. Schools that receive an "A" or improve a letter grade receive a bonus equivalent to $100 per child, which can be spent on anything including bonuses to teachers or all staff. The school advisory council determines how the bonus funds will be spent. In the first four years of the program, 140 (1997-98), 319 (1998-99), 997 (1999-2000), and 1197 (2000-01) schools received recognition awards equivalent to $100 per student in the school. Among the 2000-01 winners, 568 were "A" schools, and 629 were schools that improved by one letter grade (Florida Department of Education, 2000; 2001).

The state's financial incentive program is accompanied by extensive data-based assistance. The state has developed a sophisticated Management Information System that can track student achievement at the district, school, classroom, and individual student level. Every student in the assessment grade levels (about 1.7 million in the state) receives a report card annually with information about his or her performance on the state assessment.

School districts receive data tapes, and schools are required to develop a data-based school performance improvement plan every year. The state also provides resources for schools that seek assistance, including opportunities to submit proposals for grant funds, librarian assistance from four librarians that provide free searches to schools needing resources to guide them in modifying curricular approaches. The librarians find scientific, research-based programs to direct schools toward research-driven decisions about curriculum and instruction.

The program tends to be more market-based than the programs in Kentucky and North Carolina. Schools designated with an "F" grade are audited, and a report is submitted to the State Education Commission. Schools are given assistance in developing a school improvement plan, they participate in an education summit, and they have priority access to available financial resources to implement their improvement plan.

In the first year of the program, there were 78 "F" schools. In 2000-01, there were no "F" schools. The calculation was modified for 2001-02 to include a measure of the performance of the students in the lowest quintile of performance in each school. As a result, 68 schools received an "F" designation this year. Under the program, students in schools that receive an "F" grade for two out of four consecutive years are eligible for a voucher that they can use to attend private schools, or public schools within or outside their district that have at least a "C" designation. Students are eligible to receive the vouchers for as long as they would have attended the "F" school. This year, 9 schools became subject to these "Opportunity Scholarship" vouchers after receiving an "F" designation for the second time (Florida Department of Education, 2002).

The program is best characterized as strong accountability. Public understanding of the grading system is weak, and significant public pressure has come to bear on schools as a result of the grading system. Anecdotal evidence suggests that families have moved to avoid attending lower performing schools, and property values have shifted to reflect demand based on school performance (Figlio & Lucas, 2001).

Iowa

Iowa is currently the only state that has adopted a pay system based primarily on the demonstration of teacher knowledge and skills. Currently, the system is voluntary, but it will be mandatory for all teachers beginning in 2003. The system provides a minimum starting salary of $28,000 for new teachers (increased from about $23,000 on average currently), mentorship of new teachers, and pay increases linked to teacher development (EdSource, 2002). The system was designed to address problems related to projected teacher shortages caused by low teacher pay relative to surrounding states, inadequate professional growth opportunities for teachers in rural areas of the state, and a significant wage gap leading to a shortage of teachers in the rural areas of the state compared to urban centers.

The plan provides for the development of teacher standards linked to the INTASC and National Board for Professional Teaching Standards, and a staged licensure system

with state mandated minimum pay levels associated with teacher licensure at the Provisional, Professional - Career I, Professional - Career II, Advanced, and Advanced - National Board Certified levels. The specific procedures for advancing from one level to the next are still in the formative stages, but currently the plan is as follows. Teachers at the Provisional level participate in a mandatory Induction Program, and teachers at the Career 1 and Career 2 levels participate in a Career Development Program. Teachers will be promoted to Career 2 after 5 years in Career I and completion of the Career Development Program. Teachers can stay at this level for the remainder of their careers, or move to the Advanced category through an evaluation process. Districts may negotiate intermediate pay advancement within each of these categories, and may adopt school-based performance pay plans to provide bonuses of up to 15% of base salary when schools meet locally negotiated performance growth targets, with actual bonus varying by teacher category (provisional, career, advanced). This year, the mentoring/induction part of the program is fully operational, with 386 of 389 districts participating in the program. In addition, thus far 13 school districts have adopted locally negotiated SBPA plans.

Teachers will also be able to receive additional salary supplements for obtaining a master's degree in the primary teaching assignment, a professional license in a second teaching area (at least Career I level), or a 5% salary differential for teaching in a shortage area (e.g., math, science or special education). Teachers achieving Board Certification continue to be eligible for a $2500 pay bonus (Iowa Pay Plan, 2000).

The political negotiations that led to agreement of the plan included commitment of significant new resources to education to provide teachers with pay increases in transition to the new plan. The state's budget situation deteriorated over the course of legislative adoption, and the final plan adopted provided for phased adoption, attenuated pay increments, and local evaluation of teacher knowledge and skills by principals. Still, the state has promised $40 million in new money to raise teacher salaries in the first year (2001-2002), and another $40 million for the second year of the program (2002-2003) (Bolton, 2002). Because the plan as adopted required that evaluators be trained and certified by the state, the state Department is in the process of developing a training plan for principals to enable them to be certified as evaluators.

The plan is under fire by teachers because the state has not identified a stable or sufficient funding source (the first year of the plan was funded by Tobacco settlement money). Since the plan was not fully funded, a limited amount of money was distributed to districts in 2001-02 to be used to raise beginning teacher salaries to $28,000, and to distribute the remaining money as they saw fit. In most districts, there was little money left over to distribute to other teachers.

In December, 2001, the state released Draft teacher standards, which include the following:

1. Demonstrates ability to enhance academic performance and support for implementation of the school district student achievement goals.
2. Demonstrates competence in content knowledge appropriate to the teaching position.

3. Demonstrates competence in planning and preparing for instruction.
4. Uses strategies to deliver instruction that meet the multiple learning needs of students.
5. Uses a variety of methods to monitor student learning.
6. Demonstrates competence in classroom management.
7. Engages in professional growth.
8. Fulfills professional responsibilities established by the school district (Iowa Department of Education, 2001).

Ultimately, the teacher standards will be used as the basis for evaluating teacher performance for advancement on the salary schedule.

Discussion

Financial incentives for teachers may provide important leverage and cohesion to a statewide comprehensive education reform strategy. Compensation is an important vehicle for reform because it captures the attention of teachers, administrators, policymakers, and the public, and fosters common understandings and attention to key educational goals.

Many states have considered implementing some combination of rewards for performance improvement, and sanctions for extremely low performance and failure to improve. Research on systems that are focused on sanctions and assistance alone suggest that sanction policies may establish low-level targets that affect a small subset of schools (those in danger of being sanctioned). Furthermore, depending on program design, they may create low-level targets so that school resources are focused largely on hitting or slightly exceeding the target. Schools whose performance is sufficiently above the sanction target are largely unaffected by the sanction and assistance policies; and schools whose performance is below the target are likely to focus narrowly on meeting the established goal to avoid sanctions (Gross, 2002). SBPA programs, in conjunction with a sanction and assistance policy, can add value by establishing improvement goals for all schools, and setting higher targets than mere sanction-avoidance.

Thus, SBPA programs enhance opportunities for leveraging change in entire organizational systems in ways that sanction and assistance policies cannot do alone. They also provide a carrot to counter the stick of sanction policies, providing opportunities for the enhancement of school culture even as sanction policies put more negative pressure on organizational systems.

For financial incentives to be effective, they must be carefully designed with attention to key elements of compensation design as it relates to the design of a comprehensive education reform strategy. The following list identifies some of these key design elements:

1. *Pay and Other Incentives*. The pay incentive should be large enough to be meaningful, but not so large as to foster gaming the system, inattention to important but unmeasured educational outcomes, or to raise questions of fairness due to

imperfect measurement of educational processes and outcomes. A bonus of $2000 to $3000 seems to be sufficient to motivate teachers without causing undue warping of educational practices and behaviors. The pay incentive provides guidance to teachers about what policymakers believe is most important, and provides them with an opportunity to sift and winnow through the myriad pressures and influences that affect their decisions about what to emphasize and how to develop their teaching practice.

The pay incentive is one element of a system of rewards that teachers associate with improving student performance under SBPA programs. Other positive outcomes that motivate teachers include intrinsic rewards associated with seeing student performance improve, opportunities for problem solving and collaborating with colleagues, satisfaction associated with personal professional growth and development, recognition that teachers are doing a good job, etc. The STAR schools program in Kentucky is a good example of recognition and professional development as schools that have sustained improvement are both recognized and tapped for their expertise in school improvement. School leaders and policymakers should be aware of the broader set of rewards and design opportunities for teachers to experience them.

2. *Performance Goals.* Goal setting is a critical feature of pay for performance systems. Goals should be set at achievable levels, and should be focused on rewarding *improvements* in performance. The goals may reward overall gains in achievement, and may target gains for particular subgroups, to ensure that these groups receive focused attention by teachers and school systems. These goals might include: reducing gaps in achievement between higher and lower SES groups; raising the performance of students at the low end of the performance distribution; increasing the representation of students from disadvantaged groups in challenging classes while reducing drop-out rates and absences, and/or meeting specific performance gains for low-performing groups.

   In establishing goals, it is important to recognize the statistical challenges inherent in setting targets for small subgroups of students. The smaller the subgroup, the more likely that year-to-year variation in performance will be due to statistical "noise" such as random variation in the abilities of students from year to year. The smaller the school, the bigger this problem becomes. Kane & Staiger (2001) suggest procedures for more accurate estimation of score gains attributable to school effects.

3. *Measures of Performance and Performance Improvement.* A significant literature has developed on measuring student learning and methods for minimizing measurement error. Some argue that following the learning progress of individual students over time is the only viable approach to measuring improvements in performance, even though this approach can be very expensive to implement (Linn & Haug, 2002). North Carolina and Kentucky have both developed systems that attempt to reduce measurement error and have invested considerable effort in understanding the trade-offs associated with various approaches (e.g., Kentucky combines performance across

grade levels to enhance reliability of measures; North Carolina uses statistical predictors of expected future performance given current performance and characteristics, controlling statistically for measurement error).

In designing measures of performance and performance improvement, policymakers need to weigh the cost of particular measurement designs against the ability of the system to provide meaningful feedback to teachers (and thereby promote perceived fairness and legal defensibility). Having reliable measures of performance improvement is important to the extent that teachers need to see a connection between their own hard work and student performance. If the measure of performance improvement is not good enough to capture this, teachers will lose faith that their effort will lead to performance gains, and will stop exerting effort to improve performance. Alternatively, unreliable measures of performance improvement may lead teachers to believe that ineffective modifications to teaching practice are in fact, effective. Thus, the measure needs to be sufficiently stable and reliable so that teachers will have confidence in their ability to improve student performance, and will make valid inferences about the efficacy of their intervention efforts.

4. *The Assessment*. Accountability systems that focus teacher effort on improving student learning on a particular assessment will lead teachers to develop strategies to improve student performance on that assessment. Thus, the assessment itself must be worthy of being "taught to." Ideally, the assessment will include sufficiently high levels of cognitive demand to move teachers toward higher levels of instruction, rather than reducing curriculum and instruction to drill and practice for low levels of cognitive demand and student learning.

5. *Standards*. A clear and specific set of academic expectations or standards should be available to all educators and the public. In addition, teachers should be trained in the standards, familiar with them, and able to articulate the relationship between the standards and their own teaching practice. Schools and school systems should be able to articulate the alignment between the overall curriculum and the state standards. The state assessment should be aligned with the standards.

6. *Materials*. Curriculum guides, sample lessons, textbooks, and other materials should be available to teachers to facilitate their ability to implement the standards in their teaching practice. A recent study of reform implementation in Illinois found that a set of principals identified as effective implementers of the state standards were for the most part unable to even identify a textbook on the market that was aligned with the state standards (Bloom, 2002). To the extent that such curriculum resources are available, the state and/or districts should be able to share that information with teachers and administrators.

7. *Data Infrastructure*. The state should have sufficient data infrastructure in place that meaningful, timely data on student performance at the classroom level should be available to schools to assist them in analyzing student performance and developing

strategies to improve future performance. This information needs to be available to teachers prior to the start of the next academic year to enable them to incorporate it into curriculum planning and realignment activities.

The data could also include information about the strategies and approaches used by high performing schools, including historically low performing or high poverty schools that have successfully improved academic performance. The Kentucky and North Carolina departments of education have developed elaborate websites and other avenues for teachers to learn about the strategies and performances of other successful schools.

8. *Leadership*. Leadership continues to emerge as an important variable separating successful from unsuccessful schools. Formal networks that provide opportunities for sharing of information among school leaders, training opportunities, and external assistance providers who can provide a fresh look at the structures, strategies and patterns of behavior within schools are essential to enhancing leadership skills for low performing schools. Principals need training in organizational change dynamics as well as instructional leadership skills to be effective change agents.

9. *Technical Assistance and Professional Development*. Most financial incentive programs also provide high stakes assistance strategies for schools that fail to improve performance. Kentucky chose to identify Highly Skilled Educators selected from the pool of Kentucky educators to provide external assistance to schools that failed to improve. The state carefully selected and trained these individuals to provide organizational and instructional leadership to guide schools to utilize successful improvement strategies. This approach appears to have the advantage of a coordinated effort at external assistance, by individuals who understood the unique regional cultures and organizational constraints facing Kentucky educators.

Most programs also provide additional resources to low performing schools to provide opportunities for them to purchase critical resources to aid their improvement, such as textbooks and curricular materials, professional development, travel to study other successful schools, and additional staff to meet important localized learning needs. These resources should be provided with clear objectives for their use, and a strategy for ongoing improvement after the resources are no longer available.

Schools need opportunities to institutionalize change processes begun through external assistance. For example, in a study of the Distinguished Educator program in Kentucky, Wakelyn (2002) found that initially low-performing high schools that were able to sustain improvement over time created a Distinguished Educator-like position in the school to sustain change efforts.

10. *Public Engagement*. The public needs to understand the accountability system in order to respond with constructive pressure to improve. This can only be done if parents and the public understand the accountability system and appropriately

interpret the performance designations. For example, while there is widespread agreement that performance pay systems should reward *improvements* in student performance rather than high *levels* of student performance, systems that are based on improvement only often have had difficulty explaining to parents why a school that routinely produces Ivy League scholars is "in decline" when a very low performing but improving inner-city school is "successful."

Kentucky is a model example of state that has developed a significant ongoing infrastructure to keep the public informed and interested in educational reform, with the Prichard Committee, a non-profit organization funded largely by private foundations, a centerpiece of the Kentucky public engagement effort. Ideally, performance pay plans should invoke public pride in education, and support for improvement efforts. Parents need to engage with struggling schools to exert pressure and provide support for their improvement. Systems that merely promote public outcry and abandonment of public education do little to facilitate school improvement. Policymakers should take care to use language symbolically, and to celebrate school successes to foster the kind of supportive public pressure needed to enable and encourage school improvement.

Low Performing Schools

A key goal of education reform is to enhance the performance of currently low performing schools, which disproportionately include schools that serve low income, minority, and non-English speaking students. Research on schools serving at-risk populations suggests that schools that succeed in overcoming risk factors and produce high levels of student achievement share the following characteristics. They:

- Use state standards extensively to design curriculum and instruction, assess student work, and evaluate teachers;
- Increase instructional time in reading and math in order to help students meet standards;
- Devote a larger proportion of funds to professional development focused on changing instructional practice;
- Implement comprehensive systems to monitor individual student progress and provide extra support to students as soon as it's needed;
- Focus their efforts to involve parents on helping students meet standards;
- Have state or district accountability systems in place that have real consequences for adults in the schools (Education Trust, 1999, p. 2-3)

These factors are consistent with the characteristics emerging from schools under school-based performance award programs designed in conjunction with a comprehensive educational reform and assistance strategy to support school change and improvement. Carefully designed financial incentive and assistance programs can provide the pressure and support low performing schools need. Many of these schools lack the organizational infrastructure to support a significant change effort. Reform strategies should therefore be designed to include opportunities for external experts with leverage to gain the attention and access to teachers and administrators, with financial

resources needed to support needed reform within the school context. Kentucky has had significant success identifying and training educators to form a cadre of external experts who know and understand organizational change processes, curriculum and instructional practices, state accountability and standards as well as local context and culture. These experts can work with schools to focus new and existing resources to move beyond the constraints that are keep them from improving, to find workable strategies that enable them to succeed in overcoming the risk factors their students face.

A state's commitment to providing the leadership, technical assistance, professional development, materials and resources necessary for low performing schools to succeed is critical to the success of financial incentives supporting school improvement in low performing schools. In a study of North and South Carolina performance incentive programs, Ladd and Walsh (2002) found that schools with large concentrations of disadvantaged students were somewhat less likely to succeed. They suggest that unless the state provides assistance to help these schools overcome their relative disadvantages, over time, lower success rates in low performing schools may discourage strong teachers from teaching in these schools. To date, however, there is no research evidence to suggest that this is a significant problem. Ladd and Walsh indicate that with this research, they "hope to highlight how important it is for states or districts to supplement any test-based accountability systems with other policies explicitly designed to improve the outcomes of students in schools with large concentrations of low-performing students" (p.16).

Resources to Aid in System Design

As state experience with financial incentive programs have grown, a growing body of literature has developed to guide policymakers in the design and implementation of standards-based accountability programs, such as school-based performance pay. Three useful checklists were developed in conjunction with work commissioned by the Council of Chief State School Officers, the National Center for Research on Evaluation, Standards, and Student Testing, and the Consortium for Policy Research in Education.

First, in an excellent document highlighting and explaining the technical challenges of using assessment for accountability, Gong and associates identify 10 key questions that frame the design decision process for accountability design (see Appendix A) (Gong & ASR SCASS, 2002, p. 3). This checklist, and the information contained in the full document reflect growing levels of experience and expertise with the use of assessment for accountability. Clearer answers are still needed as to what levels of reliability, validity, and sensitivity teachers, the legal system, educators, and the public will deem sufficient to meet defensibility standards and provide accurate information regarding student learning gains and unmet learning needs. Ongoing investment in reliability analysis is critical to legal and professional defensibility, but is often overlooked by policymakers and system administrators (Gong & ASR SCASS, 2002).

A number of groups have also identified standards for the development of state assessment systems, and particularly those designed for accountability purposes. The National Center for Research on Evaluation, Standards, and Student Testing (CRESST),

in partnership with the Consortium for Policy Research in Education (CPRE) and the Education Commission of the States (ECS) have developed "Standards for Educational Accountability Systems" which also provide important and useful guidance to policymakers interested in effective accountability design (Baker, Linn, Herman & Koretz, 2002). These standards are reproduced in Appendix B.

A third checklist for the development of school-based performance pay systems developed by Odden & Kelley includes key design principles for the development of performance awards for teachers. These principles were designed in a two-year process working with national teacher leaders identified and selected by the National Education Association, the American Federation of Teachers, and the National Board for Professional Teaching Standards (See Appendix C).

Recommendations

The State of Washington has strong interest in developing an effective accountability and support system that will lead to long-term improvement in student performance, particularly in high poverty, low-performing schools (Brooks, 2000). School-based performance pay could be an important element of an educational reform strategy to leverage, focus, and reward teacher efforts to improve student learning and performance. To be effective, such a system needs to be designed in conjunction with a comprehensive strategic commitment by the state to provide the resources necessary to develop teacher and school capacity to improve student performance, particularly in schools serving large concentrations of disadvantaged students.

The following recommendations grow from the research reviewed above, and provide a process for designing and implementing a financial incentive plan to enhance school performance:

1. *Develop a school-based performance pay plan linked to Washington State Standards and to the goals of the No Child Left Behind Legislation designed as part of a comprehensive education reform strategy to improve schools*.

The specific features of the performance pay plan, along with the system of related policy efforts to enhance student performance (particularly in low- performing schools) should be designed with input and guidance from teachers and other stakeholders, as well as technical experts who have experience with test-based accountability systems, and should include the following features:

- The plan should be based substantially on improvements in academic performance, but should also include other measures of school performance, such as graduation rates, student attendance, and measures of transition post high school.
- Goals should be established at reasonable levels, so approximately 20 to 30% of schools may be eligible for awards in any one year.
- The bonus should be set at $2000-$3000 per teacher, or no more than about 5% of the average teacher salary. Pay bonuses may also be awarded to principals, classified staff, and districts achieving high levels of performance improvement

across schools. Plans that have included these bonuses typically pay the same amount to principals and district administrators as to teachers, and about half that amount for classified staff.

Teacher involvement in system design is critical so that teachers see a clear line of site between their effort and student performance, so that they perceive that the system is fair, and so that their expertise can be tapped to avoid problems in design that will lead to undesirable outcomes.

2. Invest in the development of intervention assistance for low performing schools that fail to improve.

Intervention should provide long-term (1-2 years) on-site assistance by educational experts who are instructional leaders trained in organizational change processes; curriculum and instruction; the intricacies of the state standards, assessment and accountability system; proven approaches to addressing the learning needs of special populations of students; and an insider's understanding of the culture and constraints facing educators in the school being assisted. The intervenors should have opportunities for professional development, and interaction with one another to share successful strategies, review current research, and identify educational resources and professional development opportunities that could be beneficial in assisting schools.

Principals should have opportunities for meaningful professional development as well, to provide them with training in instructional leadership and organizational change processes, and to motivate them to provide leadership in school improvement.

Successful schools serving similar populations of students should be used as potential models for low performing schools.

3. Consider the systemic relationship between preparation, certification and licensure, National Board Certification, professional development, and the skills needed to enhance student performance on the state assessment.

Work to align educator and administrator preparation, certification and licensure, and professional growth so that they support knowledge and skill development for student learning related to the state's key educational goals. For both educators and administrators, this includes training in content-specific pedagogy; and for administrators this means training in instructional leadership, organizational change processes, and data-based decision-making.

4. Engage in a concerted, ongoing effort to communicate educational reform efforts, standards, assessment and accountability to teachers, school districts, and the public.

Carefully consider labels used in the accountability system, and work to communicate fully to parents the meaning of performance categories, and how to interpret school performance based on both the *level* of performance and *improvements* in performance.

Work to communicate to parents what role they can play in strengthening the education system, and meeting the learning needs of their individual child.

5. *Incorporate a meaningful evaluation system into the program design, including key outcomes identified by CRESST.*

Specifically, determine the degree to which the system:

- builds capacity of staff;
- affects resource allocation;
- supports high-quality instruction;
- promotes student equity access to education;
- minimizes corruption;
- affects teacher quality, recruitment, and retention; and
- produces unanticipated outcomes.

The state should invest sufficient resources on an ongoing basis in an effort to refine the system to address problems as they emerge.

References

Baker, E.L., Linn, R.L., Herman, J.L, & Koretz, D. (2002). Standards for Educational Accountability Systems. The CRESST Line, Winter, p. 1-4.

Bloom, T.P. (2002). Principal Behaviors Associated with the Implementation of State Academic Standards. Unpublished Doctoral Dissertation, University of Wisconsin-Madison.

Bolton, K.A. (2002). Shortcut in teacher licensing proposed. Des Moines Register, February 3, 2002. http://desmoinesregister.com/new/stories. Obtained May 10, 2002.

Brooks, S.R. (2000). How States Can Hold Schools Accountable: The Strong Schools Model of Standards-Based Reform. Seattle, WA: Center on Reinventing Public Education.

California Department of Education. (2002). Public Schools Accountability Act Awards: 2001-02 Program Comparisons. http://www.cde.ca.gov/psaa/awards/compare01.htm. Obtained May 10, 2002.

Chan, Y., Galarza, G., Llamas, S., Kellor, E., & Odden, A. (1999). A case study of the Vaughn Next Century Learning Center's school-based performance award program. http://www.wcer.wisc.edu/cpre (January 4, 2001).

Clotfelter, C.T. & Ladd, H.F. (1996). Recognizing and Rewarding Success in Public Schools. In H.F. Ladd (Ed.), Holding Schools Accountable: Performance-Based Reform in Education. Washington, D.C.: Brookings Institution.

EdSource. (2002). Teacher Pay in California: Is it fair? Is it competitive? Is it enough? Palo Alto, CA: author.

Education Commission of the States. (2002). State Performance Indicators: January 2002. Denver, CO: author. Paper accessed September 10, 2002 at http://www.ecs.org/clearinghouse/32/12/3212.htm.

Education Trust. (1999). Dispelling the Myth: High Poverty Schools Exceeding Expectations. Report of the Education Trust in Cooperation with the Council of Chief State School Officers and partially funded by the U.S. Department of Education. Washington, D.C.: author.

Elmore, R. F. (2002). Bridging the Gap Between Standards and Achievement: The Imperative for Professional Development in Education. Washington, D.C.: Albert Shanker Institute.

Figlio, D. N. & Lucas, M.E. (2001). What's in a grade? School report cards and house prices. Paper presented at the American Educational Research Association

Conference, Seattle, WA. Paper accessed June 12, 2002 at http://www.nber.org/~confer/2001/si2001/figlio.pdf.

Florida Department of Education (2002). Florida School Grades, 2001-2002. Tallahassee, FL: author. Accessed June 12, 2002 at http://www.firn.edu/doe/schoolgrades/.

Florida Department of Education. (2001). Crist releases school grades. Press release dated May 30, 2001. Accessed July 30, 2002 at http://www.firn.edu/doe/bin00031/releases01/010530.htm.

Florida Department of Education. (2000). 997 schools receive recognition awards. Press release dated September 6, 2000. Accessed July 30, 2002 at http://www.firn.edu/doe/bin00031/releases_tg/000906.htm.

Gong, B. & ASR SCASS. (2002). Designing School Accountability Systems: Towards a Framework and Process. The State Collaborative on Assessment and Student Standards (SCASS) Accountbility Systems and Reporting (ASR) Consortium. Council of Chief State School Officers, Washington, D.C.

Gross, B. (2002). Accountability and the search for instructional strategies: Changes in search activity by elementary teachers in the Chicago Public Schools after the implementation of outcomes based accountability. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Hall, G.E. & Caffarella, E.P. (1997). Third year implementation assessment of the Douglas County, Colorado School District Performance Pay Plan for teachers. Douglas County School District, Castle Rock, CO.

Hatry, H.P., Greiner, J.M., & Ashford, B.G. (1994). Issues and case studies in teacher incentive plans (2nd ed.). Washington, D.C.: Urban Institute Press.

Heneman, H.G. III. (1998). Assessment of motivational reactions to teachers to a school-based performance award program. Journal of Personnel Evaluation in Education, 2(2), 43-59.

Heneman, H.G. III & Milanowski, A. (1999). Teachers attitudes about teacher bonuses under school-based performance award programs. Journal of Personnel Evaluation in Education, 12(4), 327-342.

Iowa Department of Education. (2001). Iowa Teaching Standards Draft Criteria. December 10, 2001. http://www.state.ia.us/educate/ecese Obtained May 10, 2002.

Iowa Pay Plan. (2000). A Proposal for a New Teacher Compensation System Based on Teacher Performance, Career Development and the National Labor Market. Report to Iowans. November, 27, 2000. Des Moines, IA.

Johnson, H., Leak, E., Williamson, G., Kellor, E., Milanowski, T., Odden A., & Hanna, J. (1999). A Case Study of the State of North Carolina's School-Based Performance Award Program. http://www.wcer.wisc.edu/cpre/tcomp obtained May 8, 2002.

Kane, T.J. & Staiger, D.O. (2001). Improving School Accountability Measures. National Bureau of Economic Research Working Paper 8156. Cambridge, MA. Paper available at http://www.nber.org/papers/w8156.

Kannapel, P.J., Coe, P., Aagaard, L., Moore, B.D., & Reeves, C.A. (2000). Teacher response to rewards and sanctions: Effects of and reactions to Kentucky's high-stakes accountability program. In B.C. Whitford & K. Jones (Eds.), Accountability, assessment and teacher commitment: Lessons from Kentucky's reform efforts. New York: SUNY Press.

Kelley, C. (1996). Implementing teacher compensation reform in public schools: Lessons from the field. The Journal of School Business Management, 8(1), 37-54.

Kelley, C. (1998). The Kentucky school-based performance award program: School-level effects. Educational Policy, 12(3), 305-324.

Kelley, C. (1999). The effects of organizational context on teacher expectancy. Unpublished manuscript. Consortium for Policy Research in Education, University of Wisconsin-Madison.

Kelley, C., Conley, S. & Kimball, S. (2000). Payment for results: The effects of the Kentucky and Maryland group-based performance award programs. Peabody Journal of Education, 75(4), 159-199.

Kelley, C., Heneman, H. III, & Milanowski, A. (2002). Teacher motivation and school-based performance awards. Educational Administration Quarterly, 38(3), 372-401.

Kelley, C. & Odden, A. (1995). Reinventing teacher compensation systems. CPRE Finance Briefs, Consortium for Policy Research in Education, University of Wisconsin-Madison.

Kelley, C., Odden, A., Milanowski, A., & Heneman, H. III. (2000). The Motivational Effects of School-Based Performance Awards. CPRE Policy Brief, RB-29-February 2000. Consortium for Policy Research in Education, University of Pennsylvania.

Kelley, C. & Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. Educational Administration Quarterly, 33(4), 474-505.

Kentucky Department of Education. (2000). Rewards for academic progress go to 618 Kentucky schools. Kentucky Teacher (October). Accessed on July 30, 2002 at http://www.kde.state.ky.us/comm/pubinfo/kentucky_teacher/oct00kyt/pg4.asp.

King, R.A. & Mathers, J.K. (1997). Improving schools through performance-based accountability and financial rewards. Journal of Educational Finance, 23, 147-176.

Klein, S.P., Hamilton, L.S., McCaffrey, D.F., & Stecher, B.M. (2000). What do test scores in Texas tell us? Issue Paper, Rand Corporation, Santa Monica, CA.

Koretz, D.M. & Barron, S.I. (1998). The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS). Santa Monica, CA: Rand Corporation.

Ladd, H. (Ed.) (1996). Holding schools accountable: Performance-based reform in education. Washington, D.C.: Brookings Institution.

Ladd, H. (1999). The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes. Economics of Education Review, 18, 1-16.

Ladd, H.F. & Walsh, R.P. (2002). Implementing value-added measures of school effectiveness: getting the incentives right. Economics of Education Review, 21, 1-17.

Legislative Analyst's Office. (2002). Accountability and Low Performing Schools, from the LAO Analysis of the 2002-03 Budget Bill (Education), pp. 109-129. Sacramento, CA: author. Accessed on May 29, 2002 at http://www.lao.ca.gov/analysis_2002/education/ed_anl02.pdf.

Linn, R.L. & Haug, C. (2002). Stability of School-Building Accountability Scores and Gains. Educational Evaluation and Policy Analysis, 24(1), 29-36.

Lawler, E.E. III. (1990). Strategic Pay. San Francisco: Jossey-Bass.

Leithwood, K. (2000). School leadership in an era of external accountability initiatives. Paper presented at the University Council for Educational Administration research conference, Albuquerque, New Mexico, November 4, 2000.

Milanowski, A. (1999). Measurement error or meaningful change? The consistency of school achievement in two school-based performance award programs. Journal of Personnel Evaluation in Education, 12(4), 343-364.

Mohrman, A., Mohrman, S.A. & Odden, A. (1996). Aligning teacher compensation with systemic school reform: Skill-based pay and group performance awards. Educational Evaluation and Policy Analysis, 18(1), 51-71.

Murnane, R.J. & Cohen, D.K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. Harvard Education Review, 56, 1-17.

North Carolina Department of Public Instruction. (2002). History of the ABCs Program. www.dpi.state.nc.us/abcs/ABCsHist.html. Obtained May 8, 2002.

North Carolina Department of Public Instruction. (2001). ABCs results highlight expansion of high performing schools; fewer schools met growth targets. Accessed July 30, 2002 at http://www.ncpublicschools.org/abc_results/results_01/press_release.html.

North Carolina Department of Public Instruction (2001a). The North Carolina State Testing Results. Preliminary State-Level Data Only. Accessed September 10, 2002 at http://www.ncpublicschools.org/Accountability/Testing/reports/green/01PrelimGB.pdf

Odden, A. & Conley, S. (1992). Restructuring teacher compensation systems. In A. Odden (Ed.), Rethinking school finance: An agenda for the 1990s (pp. 41-96). San Francisco: Jossey-Bass.

Odden, A. & Kelley, C. (2002). Paying teachers for what they know and do: New and smarter compensation strategies to improve schools (Second Edition). Thousand Oaks, CA: Corwin Press.

Okamoto, L. (2001). Pay-plan inequity raises teachers' ire. Des Moines Register. December 10, 2001. http://desmoinesregister.com/news/stories/c4780927/16840489.html. Obtained May 10, 2002.

Olson, L. (2002). Calif. Board picks ETS to run testing program. Education Week, May 8, www.edweek.com. Obtained May 10, 2002.

Pankratz, R.S. & Petrosko, J.M. (Eds.). (2000). All children can learn: Lessons from the Kentucky reform experience. San Francisco: Jossey-Bass.

Poggio, J.P. (2000). Statewide performance assessment and school accountability. In R.S. Pankratz & J.M. Petrosko (Eds.), All children can learn: Lessons from the Kentucky reform experience. San Francisco: Jossey-Bass.

Prichard Committee. (2001). Kentucky School Updates: Assessment and School Accountability. Prichard Committee website, http://www.prichardcommittee.org/pubs/updates_assessment.pdf. Obtained May 9, 2002.

Salyers, F. (2001). National Certification: What's it worth? Kentucky Teacher (May), http://www.kde.state.ky.us/comm/pubinfo/Kentucky_Teacher/May01KYT/Pg3.asp Obtained May 10, 2002.

Salyers, F. (2002). Focus on Students: New education laws support state board goals. Kentucky Teacher (May), p. 1. Frankfort, KY: Kentucky Department of Education.

Schuster, J.R. & Zingheim, P. (1992). The new pay: Linking employee and organizational performance. Lexington, MA: Lexington Books.

Stecher, B.M. & Barron, S. (1999). Test-based accountability: The perverse consequences of milepost testing. Paper presented at the annual meeting of the American Educational Research Association (AERA), April 21, 1999, Montreal, Canada.

Stecher, B.M. & Hamilton, L.S. (2002). Putting theory to the test: Systems of "educational accountability" should be held accountable. Rand Review, (Spring), 17-23.

Wakelyn, D.J. (2002). Leveraging Change in Low-Performing Schools: Kentucky's Distinguished Educators. Unpublished dissertation, University of Wisconsin-Madison.

Wilhoit, G. (2001). New Performance Standards can be Mile Markers on Our Road to Academic Proficiency. Kentucky Teacher (May). Frankfort: Kentucky Department of Education.

Appendix A
Ten Questions that Frame the Design Decision Process for Accountability Design

1. What are the purposes of the accountability system?
2. What are the main contexts, political and otherwise?
3. What are the main legal and policy constraints or specifications?
4. What are the units of performance, accountability, and reporting?
5. What are schools/students (or others) to be held accountable for?
6. What accountability decisions will be made, and with what consequences?
7. How will results be reported?
8. What data are available and will be used in the accountability system?
9. How will data be combined to make an accountability judgment?
2. How will the accountability system be monitored and evaluated?

Source: Gong & ASR SCASS, 2002, p. 3

Appendix B
Standards for Educational Accountability Systems

A. Standards on System Components

7. Accountability expectations should be made public and understandable to all participants in the system.
8. Accountability systems should employ different types of data from multiple sources.
9. Accountability systems should include data elements that allow for interpretations of student, institution, and administrative performance.
10. Accountability systems should include the performance of all students, including subgroups that historically have been difficult to assess.
11. The weighting of elements in the system, different test content, and different information sources should be made explicit.
12. Rules for determining adequate progress of schools and individuals should be developed to avoid erroneous judgments attributable to fluctuations of the student population or errors in measurement.

B. Testing Standards

7. Decisions about individual students should not be made on the basis of a single test.
8. Multiple test forms should be used when there are repeated administrations of an assessment.
9. The validity of measures that have been administered as part of an accountability system should be documented for the various purposes of the system.
10. If tests are to help improve system performance, there should be information provided to document that test results are modifiable by quality instruction and student effort.
11. If test data are used as a basis of rewards or sanctions, evidence of technical quality of the measures and error rates associated with misclassification of individuals or institutions should be published.
12. Evidence of test validity for students with different language backgrounds should be made publicly available.
13. Evidence of test validity for children with disabilities should be made publicly available.
14. If tests are claimed to measure content and performance standards, analyses should document the relationship between the items and specific standards or sets of standards.

C. Stakes

15. Stakes in accountability systems (or incentives and sanctions) should apply to adults and students and be coordinated to support system goals.

16. Appeal procedures should be available to contest rewards and sanctions.
17. Stakes for results and their phase-in schedule should be made explicit at the outset of the implementation of the system.
18. Accountability systems should begin with broad, diffuse stakes and move to specific consequences for individuals and institutions as the system aligns.

D. Public Reporting Formats
    19. System results should be made broadly available to the press, with sufficient time for reasonable analysis and clear explanations of legitimate and potential illegitimate interpretations of results.
    20. Reports to districts and schools should promote appropriate interpretations and use of results by include multiple indicators of performance, error estimates and performance by subgroup.

E. Evaluation
    21. Longitudinal studies should be planned, implemented, and reported evaluating effects of the accountability program. Minimally, questions should determine the degree to which the system:
        a. builds capacity of staff;
        b. affects resource allocation;
        c. supports high-quality instruction;
        d. promotes student equity access to education;
        e. minimizes corruption;
        f. affects teacher quality, recruitment, and retention; and
        g. produces unanticipated outcomes.
    22. The validity of test-based inferences should be subject to ongoing evaluation. In particular, evaluation should address:
        a. aggregate gains in performance over time; and
        b. impact on identifiable student and personnel groups.

Source: Baker, Linn, Herman & Koretz, 2002

Appendix C
Process and Design Principles for the Performance Pay for Teachers

Process Principles

Ten key process principles are important to the successful development, design, and implementation of a new compensation system.

1. ***Involvement of all key parties***, and especially those whose compensation is being affected, is the preeminent principle for successfully changing compensation policies. Teacher unions, administrators, school boards and the public all should be centrally involved in the process of development, design and implementation.

2. ***Broad agreement on the most valued educational results*** is also crucial. All parties – teachers, administrators, board members, parents and the public – need to agree on the results that are most valued.

3. ***Sound, comprehensive evaluation systems*** need to be in place to assess teacher knowledge and skill development in a skills-based pay system, and to evaluate organizational products and processes to be rewarded through group-based performance awards. Assessment mechanisms might include measures of student achievement, parent satisfaction, and teacher and administrator skills, knowledge, and performance.

4. ***Adequate funding*** which is integrated within the school finance structure is less likely to be vulnerable to cuts than a separate funding pool. Lack of funding and a lack of a long-term funding commitment have been key aspects of the downfall of many efforts to reform compensation in education. Transition funds often are needed to move from the old to the new structure, and performance bonuses need a stable funding pool.

5. ***Investments in ongoing professional development*** are key to skills- and competency-based pay structures. Such investments should be in the range of 2-3 percent of the operating budget.

6. ***Quotas should be avoided***. All schools meeting performance improvement targets should be rewarded, not just a fixed percentage of schools. Organizational excellence is dependent on consistent rewards for improvements in performance.

7. ***General conditions of work*** must be addressed. The better the conditions of work in a school (teacher involvement in decision-making, sound facilities, availability of materials, safety, etc.), the more likely a new form of compensation can be implemented successfully. A corollary to this principle is that the compensation system should be designed with the general conditions of work in mind. For example, skills assessment in a high-involvement school should incorporate teachers fully in the assessment process.

8. ***Management maturity*** is also important. Administrators and the school board should have good working relations, and the administration should develop a history of working cooperatively with teachers and their unions to further system goals and objectives. Restructuring the salary schedule should occur in an environment characterized by interest-based bargaining, in which each party recognizes the interests and concerns of other parties.

9. ***Labor maturity*** goes hand-in-hand with the behavior of the administration. Teacher associations, and their members, need to have positive commitment to the academic goals of the school, good working relations among themselves, and a tradition of working with management toward education system key goals.

10. ***Persistence*** until the plan is "perfected" is the key to long-term success. Most plans have initial "bugs" and are viewed with skepticism by some employees. Thus, persistence is needed to continue implementation, to revise the plan when problems are identified, and to encourage full participation to see how the plan works when fully implemented (Kelley & Odden, 1995).

Design Principles

1. *Fairness*: The system must be perceived as fair, which includes attention to the validity and reliability of the assessment instrument, attention to pre-existing differences between schools and school populations that make it more or less difficult to achieve program goals;
2. *Comprehensibility*: The system must be transparent enough that teachers and administrators can figure out what changes are needed in curriculum and instruction to address existing deficiencies, and can figure out what they did right when scores improved.
3. *Incentive-behavior compatibility*: The system should be designed so that the incentives encourage desirable (rather than undesirable) behaviors. For example the assessment itself might be designed so that "teaching to the test" results in strengthening of curriculum and instruction rather than weakening it.

Regarding design of the performance award program itself, the program should be designed so that:

4. *Rewards are assigned to whole groups rather than individuals* to encourage and facilitate collaborative problem solving at the school or departmental level. This might also mean that everyone in the school – teachers, support staff and the principal are all eligible for the reward
5. *Be very clear about what performance is most valued*. Consider including multiple measures, such as student achievement across a range of subjects, student attendance, graduation rates, etc. The system will get more of what is rewarded and less of other system outcomes.

6. *Base the performance standard for each school on improvement in student performance*.
7. *Provide an integrated and protected funding pool that will enable all schools that achieve the goals receive an award.* Stability in performance award funding over time is essential for the award to serve as an effective incentive for performance improvement.
8. *Provide awards that are valued by teachers*. Average bonuses in the $2000 to $3000 range seem appropriate.
9. *If offering a salary award, provide it as a bonus, not as an addition to base pay*. The bonus should be paid as soon as practicable after the performance to maximize incentive impact.
10. *Provide teachers with professional control over their work environment.* If teachers are to be held accountable for improvements in performance, they need to have the ability to make needed changes in the school organization to achieve those performance gains (Odden & Kelley, 2002)

Source: Kelley & Odden, 1995; Odden & Kelley, 2002